It's About Time: Analog Clock Reading in the Wild

Charig Yang, Weidi Xie, Andrew Zisserman

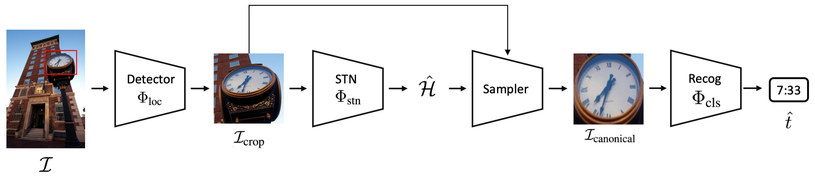
Visual Geometry Group, University of Oxford

Summary

- New task: reading analog clocks
- No labelled data available to train or evaluate on
- Steps:
  1. train using synthetic generator
  2. generate pseudo-labels on real time-lapse videos
  3. use the fact that time flows uniformly to pick reliable labels
  4. add reliable videos to training set and retrain
  5. repeat (2 – 4) until satisfied

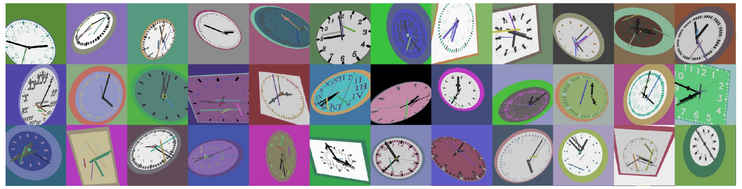
Architecture

- Simple framework: crop – align – read
  - Crop: off-the-shelf detector [1]
  - Align: direct regression for transformation matrix [2]
  - Read: direct classification for time
- but we don't have labelled data to train this with!



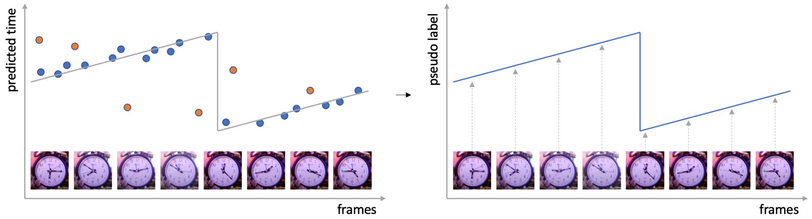
Stage 1: Synthetic data

- Train on synthetic generator. Works ok, but there is generalisation gap

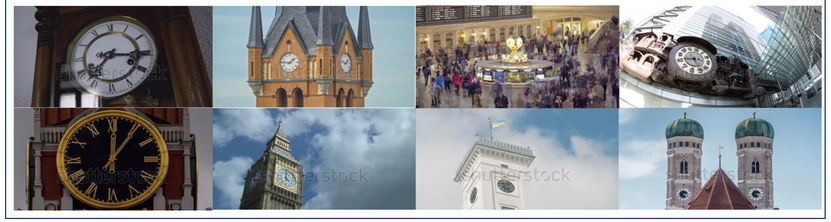


Stage 2: Iterative pseudo labelling

- Use model trained on synthetic dataset to generate pseudo labels
- Idea: we exploit **uniformity** of time, meaning it flows at a constant rate
- Fit a line* with RANSAC [3], if succeeds, add pseudo-labels to training set
  (*actually a sawtooth wave, cyclic relationship between 11:59 and 0:00)



- Timelapse dataset: 3443 unlabelled videos scraped from the internet
- Train new model, then repeat the process to get even more pseudo labels



Results

- No current benchmark for evaluation
- 3 new datasets for benchmarking, totalling 4,472 images with time labels
- Metric: both hour and minute have to be correct within +-1min
- ~72% top-1 accuracy, ~82% top-3 (all previous methods fail completely)

COCO (size: 1911) OpenImages (size: 1317) Clock Movies* (size: 1244)

(*may be subject to copyright, but will at least release the other two)

Applications

- Fun
- Time correction in metadata
- Video forensics (spotting fake videos)
- New method of searching, retrieving or grouping based on time
- Method applicable to other analog scales (scientific instruments / industrial gauges), with some adaptation

Acknowledgements

This research is supported by UK EPSRC Centre of Doctoral Training in Autonomous Intelligent Machines and Systems (AIMS), a Royal Society Research Professorship, and the UK EPSRC Programme Grant Visual AI (EP/T028572/1)

References

[1] Liang et al. "CBNetV2: A Composite Backbone Network Architecture for Object Detection". arXiv:2107.00420, 2021.

[2] Jaderberg et al. "Spatial Transformer Networks". In Proc. NeurIPS, 2015.

[3] Fischler et al. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24.6 (1981)